DOCUMENT RESUME

ED 213 729 TM 820 067

AUTHOR Conklin, Jonathan E.; Burstein, Leigh

TITLE Multilevel Approaches to the Evaluation of Educational Effectiveness. Studies in Measurement and Methodology, Work Unit 2: Analysis of Multilevel Data.

INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.

SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Nov 79
GRANT OB-NIE-G-78-0213
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Covariance; Correlation; Data Analysis; Mathematical Models; Program Effectiveness; Program Evaluation; Regression (Statistics); *Research Problems; *Statistical Analysis
IDENTIFIERS Generalizability Theory; *Unit of Analysis Problems

ABSTRACT

Educational outcomes are affected by student level, classroom level, and school level characteristics. The fact that educational data are multilevel in nature poses serious analysis questions. Though strong arguments can be made for focusing on a single level of analysis, such studies have several basic limitations: the choice of analytic level should be based on substantive, not statistical, concerns; inferences may be affected by aggregation bias; analyses at one level may automatically alternate the effects of variables from other levels; and they reveal little information about processes that exist within classrooms or within schools. This paper reviews evidence to support the claims that: distinct and important issues are addressed at each analytical level, that relationships may change at different levels, and that these changes may provide information not obtainable in any single level of analysis. Two studies are reviewed which empirically examine these issues and analyses are performed on an available database to provide specific examples of effects at multiple levels. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED213729

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DELIVERABLE - November 1979

STUDIES IN MEASUREMENT AND METHODOLOGY

Work Unit 2

Analysis of Multilevel Data

Leigh Burstein
Project Director

MULTILEVEL APPROACHES TO THE EVALUATION OF EDUCATIONAL EFFECTIVENESS

Jonathan E. Conklin, Leigh Burstein

THE USE OF WITHIN-GROUP SLOPES AS INDICATORS OF GROUP OUTCOMES

Leigh Burstein, M. David Miller, Robert L. Linn

(These two papers are in fulfillment of work proposed as "Theoretical, Conceptual, and Empirical Examination of Alternative Analytic Models for Analysis of Multilevel Data")

Grant Number
OB-NIE-G-78-0213

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.S. Gray
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 067

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California . Los Angeles

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Acknowledgements

We gratefully acknowledge the valuable assistance of Frank Capell in reviewing successive drafts of both of these papers.

The paper "Multilevel Approaches to the Evaluation of Educational Effectiveness" by Jonathan E. Conklin and Leigh Burstein was reviewed by Robert Linn and others. An early version was presented at the AERA meeting, 1979. This version will be presented at the AERA, 1980.

The paper "The Use of Within-Group Slopes as Indicators of Group Outcomes" by Leigh Burstein, M. David Miller, and Robert L. Linn was reviewed by D. Rogosa, J. E. Gustoffson, W. Honey, and L. Belandria. An early version was reported at the AERA meeting, 1979. The current version is under journal review.

This work was supported in part by a grant from the Spencer Foundation to the Graduate School of Education.

MULTILEVEL APPROACHES TO THE EVALUATION OF EDUCATIONAL EFFECTIVENESS

Jonathan E. Conklin, Leigh Burstein

Multilevel Approaches to the Evaluation of Educational Effectiveness

Jonathan E. Conklin and Leigh Burstein

Educational evaluators are often interested in variables that actually represent information at several different levels. They reason that educational outcomes are affected by student level, classroom level, as well as school level characteristics. This multilevel nature of educational data raises questions concerning the level(s) on which analyses should be focussed. Many studies measure student characteristics, classroom practices, and school policies, but average all student and classroom indices to the school level and then perform analyses on these school averages. Other studies include classroom and school variables in student level analyses. A problem of such approaches is that they attempt to draw conclusions about effects at all levels on the basis of analyses focussed on only one of them. Unfortunately, these types of crosslevel inferences may lead to grossly inaccurate and misleading statements about educational effects.

Unit of Analysis

The very fact that educational data are multilevel in nature poses serious analysis questions. Most studies view the issue as a unit of analysis problem and are therefore concerned about choosing a single level at which variables will be aggregated (or disaggregated) and analyses will be performed. Ultimately, their decision of "appropriate" unit is based on a number of concerns.

Several major studies have treated the school as the unit of analysis (e.g., Coleman et al., 1966; NORC, 1973). Their rational is that observations

within schools are highly interdependent (since students within schools share common experiences) and thus violate assumptions of independence for student and class level analyses. In addition, they see the treatment (or program) as varying across schools, but as uniform within schools. Another reason offered is that school means may be more appropriate since measures lack reliability at the student level. The use of school means also is considered practical because there is less need to be concerned about missing student level data.

Other researchers (e.g., Anderson, Evertson, and Brophy, 1978) have claimed that the class is the appropriate unit of analysis because program effects on students are most directly mediated by teachers. They argue that it is teachers that actually create the learning environment and make decisions about the use of time and materials. They also note, similar to arguments for school level analyses, that observations on students within classes are not independent and that average scores are more reliable than individual level scores.

Studies that treat the student as the unit of analysis (e.g., Stebbins et al., 1977) suggest that students should be the primary focus of interest since they are the actual recipients of the treatment. The number of students also enhances the possibility of identifying significant effects. Furthermore, with student level analyses, individualization and aptitude-treatment interactions can be effectively examined.

Though these approaches are each based on a strong rationale, they are not without their drawbacks. Studies analyzing school means are unable to identify the full effectiveness of educational processes which vary primarily within, not between schools. Studies which focus on class means cannot adequately examine possible aptitude-treatment interactions or the

effects of individualization. Student-level approaches may have technical problems because of instrument unreliability and the violation of independence assumptions.

Limitations of Single-Level Approaches

Any study which focusses on only a single level analysis has several basic limitations. First, the choice of analytical level should be strongly based on substantive concerns. Many of the arguments that are presented for choosing analytical units are concerned with such statistical issues as the independence of units, instrument reliability, sample size, and missing data. Though these are important concerns, they should not exclusively determine analysis strategies.

The choice of analysis units should give full consideration to the foucs of each research question being addressed. Hypotheses about the effects of student background, classroom instruction, or school policies on student learning are best examined at their corresponding analytical levels. No single level of focus can accurately identify effects at all three. The choice of analytical level is a decision that must be made separately for each research question by considering the level of the processes and characteristics of concern.

Second, making inferences about program effects at one level based on results obtained at another may be seriously affected by aggregation bias. Researchers have shown that correlations between variables typically change when data are aggregated (Thorndike, 1939; Robinson, 1950). That is, variables appear to have different relationships with each other depending on whether aggregate o individual level data are used. The implication is that using aggregate data to make inferences about

individuals may lead to conclusions that differ from those that would be obtained if individual level data were used. Under only the most stringent conditions will parallel analyses at each of several levels reveal the same effects (Hannan and Burstein, 1974; Burstein, 1978b). Whether or not the conditions are met, variables with effects at the individual level, when aggregated to the class level, become confounded with classroom level processes, and both are confounded with school variables when aggregated to the school level. With these kinds of confounded effects between variables at different levels, it is difficult to determine whether differences in student outcomes should be attributed to differences in background or to differences in classrooms or schools.

Third, analyses at one level may automatically attenuate the effects of variables from other levels that have been included in the analyses. For instance, if school indices are included in student level analyses, the apparent size of their effects may be diminished because they lack variability within schools. This results because every student within a given school has the same value on the school index. When student and classroom characteristics are aggregated to the school level, their apparent effects may be diminished since their variability has essentially been averaged out (i.e., they vary more within schools than between schools). On the other hand, some student level characteristics that vary more from school to school than within schools (e.g., socioeconomic status) may appear to be more effective when aggregated and included in school level analyses. Analyses performed at one level, then, have a natural tendency to give misleading indications about the effects of variables from another level.

Finally, because most of the studies of educational effects concentrate exclusively on between-school and between-student analyses, they can reveal little information about processes that exist within classrooms or within schools. Relationships found at the between group level of analysis may not exist within those groups. As an example, consider the variable of teacher experience. It is possible that schools with more experienced teachers also exhibit higher average achievement levels that schools with less experienced teachers, an effect which would be revealed from between-school analyses. It is also possible, though, that within schools and within classes, the more experienced teachers best facilitate the learning of already motivated (and higher achieving) students whereas the lower achievers may be better inspired by the enthusiasm and creativity of younger, less experienced teachers. This interaction effect between teacher experience and student ability (or motivation) could only be revealed by within-school or within-class analyses. Individual level analyses only serve to confound within-group and between-group effects (Burstein, 1978a; Cronbach, 1976; Martin, 1978). The within-group approach also avoids some of the statistical problems of individual level analyses (lack of independence) and acknowledges the fact, unlike group level analyses, that students within classes and schools react differently to specific teacher behaviors and school policies.

No single level (unit) of analysis can provide full clarification of the effects of student, class, and school characteristics on achievement outcomes. The problems of aggregation bias, the presence of distinct within- and between-group effects, and the fact that research questions are typically posed about processes at each of the levels all suggest

that a multilevel approach to analysis will result in a more complete understanding of educational effectiveness.

Purpose

The remainder of this paper briefly reviews evidence to support the claims that: distinct and important issues are addressed at each analytical level, that relationships may change at different levels, and that these changes may provide important information not obtainable in any single level of analysis. Two studies are reviewed which empirically examine these issues and analyses are performed on an available data base to provide specific examples of effects at multiple levels.

Teacher Effectiveness Study

In a reanalysis of the Anderson et al. (1978) study of teacher effectiveness, Martin (1978) directly confronted the unit of analysis issue. She compared results of class level, reading group-within-class level, and individual-within-reading-group level analyses. The basic analytical models employed were built on a multiple regression framework with reading achievement seen as a function of reading readiness, teacher behavior, and their interaction. At the class level, class means were analyzed to examine the effects of teacher behavior on class mean achievement. At the reading group-within-class level, reading group means were entered with binary vectors representing class membership (thus controlling for class level differences). Such analyses examined whether the amount of teacher behavior reading groups receive, relative to other reading groups in the same class, is related to achievement. At the individual-within-reading-group level, individual scores were entered along with binary vectors representing reading group membership (to control for group

differences). These analyses examined whether the amount of teacher behavior students receive, relative to other students in the same reading group is related to achievement.

Overall, the results suggested that the strongest effects (even after taking into account degrees of freedom differences) are at the student-within-reading-group level. Martin interpreted this as indicating that since, here, "teacher-student interactions occur in the context of a peer group ... students are in a position to directly compare their treatment and performance with that of their peers" (p. 26). Results also tended to be inconsistent across the different levels. Martin explained this by stating that "not only do analyses with different analysis units answer different questions, different patterns of results across levels of analysis reveal different types of information about teacher behaviors" (p. 127). Her results indicate that the way in which students react to teacher behaviors and learning opportunities depends to a large extent on their relative standing within the group. In addition, certain directives from the teacher may be interpreted differentially by groups and individuals differing in ability and background. Such results may have important implications about program features which emphasize the individualization of instruction and the structuring of groups (classes) and may better enable teachers to be effective with groups of differing characteristics.

IEA Study

In his analyses of data from the studies of the International Association for the Evaluation of Educational Achievement (IEA), Burstein (1978a; Burstein et al., 1978) employed multilevel methodology. The basic models examined were multiple regression models which posited achievement

to be a function of general verbal ability, indices of background conditions in the home, and measures of schooling. Analyses were conducted at the between-school, pooled within-school, and between-student levels. Comparisons of unstandardized regression coefficients at different levels for specific variables were used to identify complex relationships between background and educational performance.

Burstein posited that whereas at the individual level, measures of family background may act as indices of parental investment in a child's learning, at the school mean level, such measures act as proxies for community wealth, resource, and attitude toward education. We can see that the effects of both constructs are important and may have definite policy implications. On the one hand, strong effects of background may indicate the need for programs to stress increased parental involvement and awareness; on the other, they may emphasize the importance of community input and support.

Cross-national comparisons conducted using within and between-school analyses may serve to identify possible influences of national social and educational policies. As an example of interpretations made from cross-national comparisons, Burstein claimed that "the difference between the U.S. and Sweden in the between-school effects of background suggests that reducing the relation of school resources to community characteristics can lessen the impact of background on school-level achievement" (p. 55). In another reference (Burstein et al., 1978), he asserted that "given the opposite pattern of between-school ability effects, the differences in within-school coefficients [of verbal ability] for England and the U.S. may be attributable to the role that between-school curriculum tracking plays in an administratively decentralized educational system. In the

absence of between-school curriculum differentiation, within-school differentiations have a stronger influence on the relationship of verbal ability and science achievement ... In England the necessity for within-school tracking mechanisms is largely alleviated by the availability of alternative schools with different curricular emphases" (p. 23).

These references are cited as indication that multilevel analyses can result in interpretations and conclusions which would be impossible to make on the basis of only a single level of analysis. The interpretations referenced above seem to support a contention that multilevel analyses may not only imply certain policies at the individual or school levels, but also at the program, and, indeed, perhaps even at the national level.

An Additional Example

To investigate further the different implications resulting from analyses performed at different aggregate levels, data from the Emergency School Assistance Program (ESAP) evaluation (NORC, 1973) were examined. This data base was selected because of its numerous available indices of student and classroom characteristics and because it contains data from three levels: student, class, and school.

From the full EASP data set, a subset of nonmissing and nearly balanced data was selected. The final sample consisted of three fifth grade classes from each of 62 schools with between 10 and 15 students sampled per class (total N = 2644). Of these schools, 11 were ESAP program schools, 14 were control schools, and 37 were "supplementary" schools. For the purposes of our investigation, all 62 schools were included in analyses in which the specific aim was not the identification of program effects. Analyses of program effects per se concentrated only on the 25 experimental and control schools.

With the ESAP data, the aim was not to contrast within-group and between-group analyses. Rather, our concern was to identify for educational researchers/evaluators, and to verify for ourselves, any possible analysis complications brought about solely from the multilevel structure of educational data. In view of the issues concerning units of analysis and since evaluations vary greatly in their choice of analytical level, we felt that it would be informative to compare results of parallel analyses across various levels. Our intent, then, was not specifically to examine the reasons for performing analyses at specific levels, nor was it to identify generalizable educational effects. With the ESAP analyses, we merely hoped to examine and identify possible differing interpretations about schooling effects that one might reach based on the same analyses at different levels. If evaluation conclusions and policy implications vary with the choice of analytical level, then perhaps large-scale educational evaluations should employ a multilevel approach.

Our approach to analyzing the ESAP data was multifaceted in that it consisted of several different types of analyses and model specifications. All analyses were simple and direct:

- variance components were examined to identify sources of variability;
- analyses of covariance were used to reveal possible program effects;
- correlations were examined at each level of aggregation; and,
- a few multilevel regression models were used to contrast between- and within-group effects.

Our analyses were limited to measures of student achievement and background, teacher characteristics, and school variables. Scales were formed similarly to those used in the original ESAP evaluation (NORC, 1973), but were slightly modified on the basis of factor analytic results.

Variance Components. By decomposing the total variability of a measure into its three component sources (students, classes, and schools), we can infer the possible differential impact of processes at different levels. For this reason we estimated variance components for a number of student indices and compared them as percentages of total variability. The results are presented in Table 1.

We can see that the relative magnitudes of within-class, between-class, and between-school contributions to total variability differ from measure to measure. Generally, measures of achievement and socioeconomic background are characterized by large between-school variance components. This appears to suggest that school-to-school differences covary more with achievement variability than do average differences in classroom emphases. Such an indication may be an artifact of analyzing together ESAP program, control, and "supplementary" schools. If program allocations affect the availability of learning resources and materials, or if the effects of desegregation at the school level affect the performance of individual students, we might expect to find a large school effect on achievement variability. The variability between classes in terms of their composition and teaching approaches may be more school determined or related. Of course, some of the school effect on achievement may be spurious due to the high correlation between achievement and socioeconomic variables that occurs because schools tend to draw from their

immediate neighborhoods which vary in income, racial compositon, etc. Students and classes within schools are relatively more homogeneous on such variables than are schools themselves.

On measures of Self-Esteem and Test Anxiety, variability tends to be most related to pupil differences with little indication of class or school influence. On the other hand, variability in Attitude Toward School appears to be largely classroom related. This suggests that specific teachers and classroom situations may account for pupil perceptions of their school experience.

Our analyses of variance components indicate that influences on student characteristics and achievement are multilevel in nature. For some variables, the school effect may be greatest; for others, classroom or student effects appear to be more important. This has important implications for decisions about further analyses. To hope that analyses at only a single level could identify the strongest influences on student outcomes is naive. Analyses at one level should not be used to make inferences concerning influences at another level. Moreover, analyses at the "wrong" level may result in policy implications that are misdirected (e.g., school oriented rather than classroom oriented approaches to improve student attitudes).

Analyses of Covariance. Analyses of covariance were used to partially control for socioeconomic influences in examining possible program effects on achievement. The analyses compared the 11 program and 14 control schools in the sample. Results are presented in Table 2. Program effects on reading and science achievement appear to have remained constant across levels of analysis, but for writing and math achievement, apparent program effects vary in magnitude and significance depending on the analysis

level used. The effects become smaller (or more negative) and ware less often significant when analyzing school-level aggregated data. Furthermore, degrees of freedom differences do not fully account for this pattern.

Such an observed pattern of effects may be due to the inadequacies of covariance adjustment of school means. As Haney (1974) points out, "adjustment of posttest scores for pretest shcould be made at the pupil level prior to any aggregation to higher levels of analysis if the full effect of the adjustment is not to be lost" (p. 29). Clearly then, cross-level inferences about program effects based on single-level analyses of covariance may be highly misleading. Effects which appear significant at one level may not be at another. To make inferences about the actual size of program effects, analyses must be performed at the level at which program impact is manifested.

Correlations. The correlation between two variables may differ in magnitude at different levels of aggregation (Thorndike, 1939; Robinson, 1950). Using the ESAP data, we calculated correlations between a number of variables and achievement at the three analytical levels. These are presented in Table 3. As we can see, correlations at lower aggregate levels are usually smaller in magnitude than those at higher levels. However, there are a few interesting exceptions. The obtained correlation between Self-Esteem and achievement shows a decrease with aggregation. This may be partly explained by recalling the fact that the student component of variance for this variable far exceeds both the class and school components. That is, the correlations at the class and school levels are smaller because at those levels variability in self-esteem is greatly reduced. For some variables, correlations are significant at one level, but not at others. Correlations of achievement with Teacher

Education, Student Attention in Class, and the Teacher's Use of Principal For Advice are significant only at the school aggregate level. The Teacher's Use of Discussion In Class is significantly correlated with achievement at the class level, but not at the student or school levels.

We may conclude that by analyzing correlations at only one level, we deprive ourselves of knowledge about relationships at other levels. School-level correlations of student indices may exaggerate actual relationships, whereas for some variables, lower aggregate-level analyses may reveal stronger correlations. The choice of appropriate level may never be clear. Indeed, comparisons across levels may be more informative to decision makers.

Regression Analyses. A few regression analyses were conducted to make comparisons across levels and to examine within-level effects. The models presented in Table 4 consider achievement to be a function of several socioeconomic factors and either classroom or school-related variables.

For the most part, effect sizes and their significance depend on the level at which the analyses were performed. Teacher Education exhibits a significant positive influence on reading and math achievement only when analyzed at the individual level (reading also at the class level). Effects of Percent Urban and Amount of Funding (both school-level indices) are only significant in student-level analyses. The Teacher's Use of the Principal For Advice is related to science achievement only when school-level analyses are performed, possibly indicating that qualities of principals determine whether they are sought out for advice (or that teacher ability to teach science is school-related).

We also examined a specific type of within-class effect by creating models in which within-class variability in achievement was included as a

predictor of student achievement in the same subject matter. Results of these analyses are presented in Table 5. We posited that with greater heterogeneity of ability within a class, it would be more difficult for a teacher to meed individual students' needs than it would be in a more homogeneous class. The table indicates that within-class variability has a significant negative effect on student achievement in reading and in math, but not in writing or science. This may suggest that in reading and math, it is more difficult to meet the needs of individual students, especially when classes exhibit much variation in ability. The teacher may be forced to slow down or cover less material in classes characterized by much variability so as to ensure that all students learn some of the basics. In classes with less achievement variability, the teacher may more efficiently teach the class as a whole. In the subjects of science and writing, instruction may be more pre-planned and structured, and thus not be so sensitive to within-class variability.

Summary

Summarizing the results of these analyses, we see that regression analyses exhibit differential effects at different levels of aggregation. Patterns of coefficients are typically unstable. Inferences about program impact based upon analyses at a single level may identify some variables as being ineffective, whereas at another level, those same variables may demonstrate strong effects. Thus, comparisons across levels should be made. Policy decisions should not be based on the results obtained from only a single level of analysis. Multilevel models may reveal effect patterns which could not be discerned from a single level analyses along (Burstein, 1978a).

The analyses of the ESAP data set clearly support the types of results obtained in the Martin study (1977) and then with the IEA data, namely that level of analysis used may strongly influence the results obtained. The fact that educational data have a multilevel structure behooves us to consider multilevel approaches to analysis.

Conclusions

The presentation of a few findings from a simple multilevel analysis, and references to two extensive multilevel studies have emphasized the need for multilevel analyses and comparisons in studies of educational effectiveness. Program evaluations must not strictly adhere to a single level of analysis. This report suggests that the resulting interpretations concerning program effects and policy implications may be misleading. Furthermore, effects may be hard to find at only a single level. Emphases may be placed on innovations or programs aimed at the wrong level of implementation.

The choice of appropriate level of analysis should not be the major concern. Important questions must be answered at each level. Policy decisions are made not only at the district or state level; they are also made by school administrators and by teachers within schools. Given the decisions that administrators must make about personnel and the allocation of school resources, as well as the instructional decisions made by individual teachers, these studies should also be concerned with providing information about within-class, between-class, and within-school effects.

To plan the analyses for each research question, the various purposes of the evaluation itself must be examined. In answering questions only about whether schools in one program benefit more than schools not in the program, we may wish to focus on analyses of school level outcomes. If

our interest is only in the possible effects of program allocations on actual teaching practices and the use of materials, perhaps we should be concerned with class level analyses. If we are solely interested in investigating possible program effects on individualization and on student outcomes over and above the effects of their abilities and backgrounds, we should focus on student level analyses. Finally, if our intent is to identify within-class (school) effects, effects of relative position within class (school), or effects of within-class (school) variation in background or ability, we must resort to multilevel models which simultaneously take into account class (school) level and student level measures. Of course, the typical large-scale evaluation is interested in all of these questions. It is inappropriate, then, to limit the focus to only on level of analysis. The inferences made and the resulting policy implications with respect to the different research questions of interest depend too greatly on the choice of level(s) to be used in each analysis.

References

Anderson, L., Evertson, C., & Brophy, J. The first grade reading group study: Technical report of experimental effects and process-outcome relationships. Austin, Texas: Research and Development Center for Teacher Education, 1978.

Burstein, L. The Pole of Levels of Analysis in the Specification of Educational Effects. Paper commissioned by the Educational Finance and Productivity Center, Department of Education, University of Chicago, under contract from the National Institute of Education, May, 1978.

Burstein, L., Fischer, K., & Miller, M.D. Social Policy and School Effects: A Cross-National Comparison. Paper presented at the IX World Congress of Sociology Meeting, Uppsala, Sweden, August, 1978.

Coleman, J.S., et al. Equality of Educational Opportunity. Washington, D. C.: U.S. Government Printing Office, 1966.

Cronbach, L.J. Research on Classrooms and Schools: Formulation of Questions, Design and Analysis. Occasional Paper of the Stanford Evaluation Consortium, Stanford University, 1976.

Firebaugh, G. A rule for inferring individual-level relationships from aggregated data. American Sociological Review, 1978, 43(4), 557-572.

Haney, W. Units of Analysis Issues in the Evaluation of Project Follow Through. Unpublished report, Cambridge, Mass.: Huron Institute, 1974.

Knapp, T. The Unit of Analysis Problem in Applications of Simple Correlation Analysis to Educational Psychology. J. Ed. Stat., 1977, 2, 171-186.

Martin, J. The Unit of Analysis Problem in Teacher Effectiveness Research. Unpublished dissertation. University of Texas at Austin, December, 1978.

NORC (National Opinion Research Center). Southern Schools: An Evaluation of the Effects of the Emergency School Assistance Program and of School Desegregation. Chicago, 1973.

Robinson, W.S. Ecological Correlations and the Behavior of Individuals. American Sociological Review, 1950, 15, 351-357.

Stebbins, L.B., et al. Education as Experimentation: A Planned Variation Model (Vol. IV-A, An evaluation of Project Follow Through). Boston: ABT Associate, April, 1977.

Thorndike, E.L. On the fallacy of inputing the correlations found for groups to the individuals or smaller groups composing them, American Journal of Psychology, 1939, 52, 122-124.

Table 2. Analyses of covariance to test program effects. Unstandardized regression coefficients.

	READING			WRITING			MATH			SCIENCE		
	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.
SES	.283*	.627*	1.068*	.466*	.388	1.227	.577*	1.080*	2.872*	.347*	.619*	1.019*
SIBLINGS	-.113*	-.376*	-.367	-.185*	-.546*	-.156	-.216*	-.804*	-.263	-.145*	-.459*	-.393
RACE	-1.487*	-.927*	-.808	-2.772*	-2.340*	-2.996	-3.744*	-3.294*	-4.552*	-1.865*	-1.329*	-1.184
HOMEREAD	.525*	1.344*	1.324*	.090*	2.832*	3.421*	1.440*	2.405*	2.637	.489*	1.503*	1.678*
PROGRAM	.186	.183	.179	.451*	.408*	.296	-.146	-.137	-.305	.107	.102	.088

*$p < .05$

Table 1. Components of variance expressed as percentages of total mean variation. (Based on the formula: $\sigma_T^2 = \sigma_S^2 + \frac{\sigma_c^2}{m} + \frac{\sigma_e^2}{mn}$ where m = 3 and n = 14.3)

	SCHOOL	CLASS W/IN SCHOOL	STUDENT W/IN CLASS
READING	58.5	24.1	17.4
WRITING	66.2	16.3	17.4
MATH-COMPUT.	60.9	21.5	17.5
MATH-CONCEPTS	64.3	18.8	16.9
SCIENCE	58.6	21.2	20.2
TOTAL ACH.	64.4	21.7	13.9
READ HELP	44.0	26.9	29.1
PARENT SATIS.	43.2	20.1	36.7
RACE	78.0	8.4	13.7
SIBLINGS	67.9	8.2	24.0
SELFESTM	17.4	12.6	70.0
ATTITUDE	6.5	49.4	44.0
TEST ANX.	34.7	24.9	40.3
SES	54.0	16.0	30.0
HOMEREAD	68.3	6.0	25.7

NOTE: See Table 6 for the scale definitions corresponding to these variable names.

Table 3. Correlations of selected variables with achievement at all three levels.

	READING			WRITING			MATH			SCIENCE		
	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.
SES	.21*	.43*	.56*	.22*	.39*	.47*	.21*	.43*	.54*	.20*	.42*	.54*
SIBLINGS	-.25*	-.55*	-.66*	-.26*	-.52*	-.54*	-.25*	-.56*	-.65*	-.25*	-.52*	-.62*
RACE	-.40*	-.54*	-.59*	-.41*	-.55*	-.58*	-.43*	-.61*	-.69*	-.40*	-.57*	-.63*
HOMEREAD	.19*	.38*	.57*	.19*	.39*	.54*	.21*	.36*	.52*	.18*	.38*	.62*
SELFESTM	.33*	.30*	.21	.31*	.27*	.17	.36*	.29*	.23*	.33*	.29*	.24
TCHEDUC	.07	.17	.32*	.04	.11	.21	.06	.13	.26	.04	.11	.25
STDATTEND	.05	.10	.33*	.06	.13	.34*	.08	.16	.33*	.05	.10	.29*
DISCUSSN	.09	.20*	.24	.09	.21*	.22	.09	.19	.18	.09	.22*	.23
DISCPLIN	-.04	-.08	-.17	-.06	-.13	-.18	-.05	-.11	-.22	-.07	-.15	-.21
PADVICE	-.01	-.03	-.19	-.03	-.09	-.24	-.02	-.06	-.26	-.03	-.10	-.33*

*$p < .05$ (correlation must exceed .15 to be practically significant)

Table 4. Unstandardized regression coefficient - regressions of achievement on student, class, and school characteristics.

	READING			WRITING			MATH			SCIENCE		
(A)	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.
SES	.23*	.56*	.91*	.43*	.53	1.02	.49*	1.03*	1.63	.37*	.51*	.85*
SIBLINGS	-.12*	-.39*	-.34*	-.20*	-.51*	-.18*	-.24*	-.80*	-.51	-.14*	-.33*	-.16
RACE	-1.52*	-1.19*	-.78*	-2.60*	-2.26*	-2.07*	-3.88*	-4.08*	-3.58*	-1.85*	-1.71*	-1.34*
HOMEREAD	.43*	.70*	1.37*	.70*	1.36*	2.69*	1.17*	1.38	2.78*	.45*	.90*	2.02*
TCHEDUC	.17*	.15*	.25	.14	.11	.17	.29*	.24	.31	.10	.08	.15
RACEDIFF	-.06	-.38	-.03	.02	-.70	-.33	-.09	-.78	-.92	.10	-.25	-.27
R^2	.20	.45	.64	.21	.43	.51	.22	.49	.65	.19	.44	.65
(B)	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.
SES	.23*	.54*	.94*	.43*	.59	.99	.49*	.98	1.57	.27*	.47	.81*
SIBLINGS	-.13*	-.44*	-.39*	-.20*	-.43*	-.26	-.25*	-.90*	-.76*	-.15*	-.38*	-.24
RACE	-1.55*	-1.37*	-.75	-2.59*	-2.44*	-2.11*	-3.91*	-4.40*	-3.71*	-1.80*	-.187*	-1.35*
HOMEREAD	.44*	.62	1.34*	.70*	1.19*	2.47*	1.18*	.90	2.37*	.45*	.82*	1.76*
PADVICE	.01	-.04	-.06	-.06	-.16	-.20	-.03	-.17	-.39	-.06	-.13	-.25*
PQUAL	-.01	.03	.01	.10	.03	-.02	.08	.08	-.002	-.02	-.01	-.04
R^2	.20	.43	.62	.21	.40	.51	.22	.48	.64	.19	.45	.66

Table 4 Continued.

	READING			WRITING			MATH			SCIENCE		
(C)	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.	STD.	CL.	SCH.
SES	.24*	.56*	.98*	.44*	.58	1.07	.51*	1.04	1.81*	.28*	.48	.85*
SIBLINGS	-.12*	-.41*	-.35*	-.19*	-.51*	-.20	-.23*	-.81*	-.58	-.14*	-.34*	-.20
RACE	-1.56*	-1.48*	-.94	-2.61*	-2.58*	-2.37*	-3.93*	-4.62*	-4.22*	-1.81*	-1.99*	-1.57*
HOMEREAD	.37*	.45	1.09	.62*	1.15	2.29*	1.04*	.50	1.76	.39*	.71	1.83*
URBAN	.65*	.39	.26	.88*	.28	.34	1.41*	.93	.86	.67*	.37	.15
FUNDS	.64*	.82	.69	.83	.97	.82	.59	.88	.90	.74*	.93*	.79
R^2	.20	.45	.63	.21	.41	.51	.23	.49	.64	.20	.46	.66

*$p < .05$

Table 5. Individual level and multilevel[1] regressions.

	READING		WRITING		MATH		SCIENCE	
	INDIV.	INDIV. + $SD_{READING}$	INDIV.	INDIV. + $SD_{WRITING}$	INDIV.	INDIV. + SD_{MATH}	INDIV.	INDIV. + $SD_{SCIENCE}$
SES	.228*	.224*	.427*	.427*	.484*	.481*	.265*	.264*
SIBLINGS	-.128*	-.132*	-.204*	-.205*	-.246*	-.255*	-.146*	-.147*
RACE	-1.552*	-1.534*	-2.593*	-2.600*	-3.918*	-3.881*	-1.801*	-1.800*
HOMEREAD	.435*	.438*	.702*	.709*	1.177*	1.200*	.451*	.453*
CORRESPONDING ACH. VARIABILITY W/IN CLASS		-.342*		-.170	-.779*	-.062		

*$p < .05$

[1]These multilevel models include, with individual level background variables, the within-class standard deviations of the dependent-variable achievement measures.

Table 6. Scale definitions for variable names.

VARIABLE NAMES	SCALES
READING	Achievement score in Reading
WRITING	Achievement score in Writing
MATH-COMPUT.	Achievement score in Math. Computations
MATH-CONCEPTS	Achievement score in Math. Concepts
SCIENCE	Achievement score in Science
TOTAL ACH.	Total achievement
READHELP	Is student given special reading help?
PARENT SATIS.	Are parents satisfied with student's grades?
RACE	Race/ethnicity (1=white, w=minority)
SIBLINGS	Number of siblings
SELFESTM	Self esteem
ATTITUDE	Attitude toward school
REST ANX.	Test anxiety
SES	Socioeconomic status
HOMEREAD	Home encouragement for reading
PROGRAM	1=program school, 0=control school
TCHEDUC	Level of teacher's education
STD.ATTEND	How many students pay attention in class?
DISCUSSN	How much class time is used in discussion?
DISCPLIN	How often is discipline used in class?
PADVICE	How often does teacher go to principal for advice?
RACEDIFF	Difference between student's and teacher's race
PQUAL	Principal's quality as rated by teacher
URBAN	Percentage urban (from 1960 census)
FUNDS	Per pupil expenditure of funds

*All scales are formed in a positive direction (i.e., higher scale scores denote higher levels of underlying characteristics).